The decomposition of the higher-order homology embedding constructed from the k-Laplacian

Finding ``independent'' loops in a manifold

Yu-Chia Chen (yuchaz@uw.edu)
Electrical & Computer Engineering
University of Washington
December, 2021

Neural Information Processing Systems
NeurIPS 2021

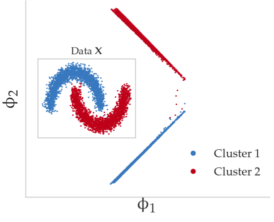
- Joint work with professor Marina Meilă
  - Department of Statistics, University of Washington.
    mmp@uw.edu

- The slides can be downloaded in
  https://bit.ly/chen_meila_21_slides

Motivation

Embedding of spectral clustering

- Structure of the embedding is known:
  - Orthogonal cone structure (OCS) [Schiebinger et al., 2015]
- Clusters (red/blue) can be identified from the embedding
- Spectral clustering := 0-homology embedding



What about the higher-order cases?

- Empirical observation [Ebli and Spreemann, 2019]
  - Embedding is a "union" of subspaces

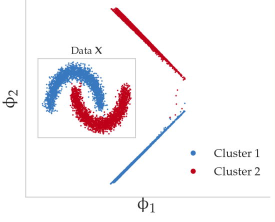
- Localize the "subcomponents" of a manifold

Main contribution

- A theoretical analysis of the above observation
  - Using the concepts of connected sum and matrix perturbation theory
- Data-driven decomposition algorithm + identifying loops (side product)

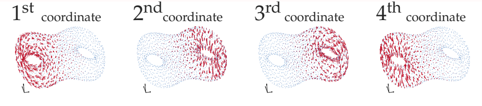
Embedding of spectral clustering

- Structure of the embedding is known:
  - Orthogonal cone structure (OCS) [Schiebinger et al., 2015]
- Clusters (red/blue) can be identified from the embedding
- Spectral clustering := 0-homology embedding



What about the higher-order cases?

- Empirical observation [Ebli and Spreemann, 2019]
  - Embedding is a "union" of subspaces
- Localize the "subcomponents" of a manifold

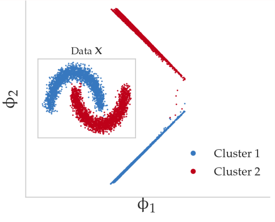


Main contribution

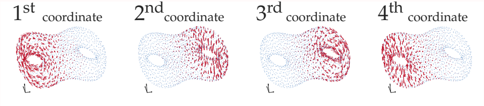
- A theoretical analysis of the above observation
  - Using the concepts of connected sum and matrix perturbation theory
- Data-driven decomposition algorithm + identifying loops (side product)

Embedding of spectral clustering

- Structure of the embedding is known:
  - Orthogonal cone structure (OCS) [Schiebinger et al., 2015]
- Clusters (red/blue) can be identified from the embedding
- Spectral clustering := 0-homology embedding



What about the higher-order cases?



- Empirical observation [Ebli and Spreemann, 2019]
  - Embedding is a "union" of subspaces
- Localize the "subcomponents" of a manifold

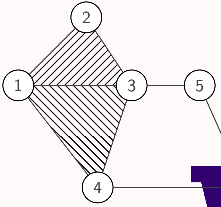
Main contribution

- A theoretical analysis of the above observation
  - Using the concepts of connected sum and matrix perturbation theory
- Data-driven decomposition algorithm + identifying loops (side product)

Introduction

(Finite samples from $\mathcal{M}$)		(Want to approximate)	
Discrete		Continuous	
Simplicial complex	SC_ℓ	Manifold	$\mathcal{M}$
k-cochain	$\boldsymbol{\omega}_k$	k-form	ζ_k
Boundary matrix	$\mathbf{B}_k$	Codifferential operator	δ_k
Coboundary matrix	$\mathbf{B}_k^\top$	Exterior derivative	d_{k-1}
Discrete k-Laplacian	$\mathcal{L}_k$	Laplace-de Rham operator	Δ_k
k-homology space	$\mathcal{H}_k \subseteq \mathbb{R}^{n_k}$	k-homology group	$H_k(\mathcal{M}, \mathbb{R})$

- Simplicial complex
  - $SC_\ell = (\Sigma_0, \Sigma_1, \cdots, \Sigma_\ell) = (V, E, T, \cdots, \Sigma_\ell)$
  - $n_k := |\Sigma_k|$
- Clique complex of G
  - fill all triangles, tetrahedrons, ..., (all k-cliques) in G



Discrete k-Hodge Laplacian and manifold geometry

(Finite samples from $\mathcal{M}$)		(Want to approximate)	
Discrete		Continuous	
Simplicial complex	SC_ℓ	Manifold	$\mathcal{M}$
k-cochain	$\boldsymbol{\omega}_k$	k-form	ζ_k
Boundary matrix	$\mathbf{B}_k$	Codifferential operator	δ_k
Coboundary matrix	$\mathbf{B}_k^\top$	Exterior derivative	d_{k-1}
Discrete k-Laplacian	$\mathcal{L}_k$	Laplace-de Rham operator	Δ_k
k-homology space	$\mathcal{H}_k \subseteq \mathbb{R}^{n_k}$	k-homology group	$H_k(\mathcal{M}, \mathbb{R})$

Symmetrized k-Laplacians [Horak and Jost, 2013]

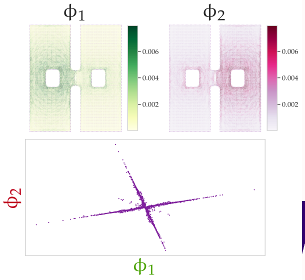
$$\mathcal{L}_k = \underbrace{\mathbf{A}_k^\top \mathbf{A}_k}_{\mathcal{L}_k^{\text{down}}} + \underbrace{\mathbf{A}_{k+1}\mathbf{A}_{k+1}^\top}_{\mathcal{L}_k^{\text{up}}}.$$

- $\mathbf{A}_\ell := \mathbf{W}_{\ell-1}^{-1/2}\mathbf{B}_\ell \mathbf{W}_\ell^{1/2}$
  - Normalized boundary matrix
- $\mathcal{L}_0 = \mathbf{A}_1\mathbf{A}_1^\top = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{K}\mathbf{D}^{-1/2}$
  - Symmetrized graph Laplacian
- $\mathcal{L}_k \in \mathbb{R}^{n_k \times n_k}$

Discrete k-Hodge Laplacian and manifold geometry

(Finite samples from $\mathcal{M}$)		(Want to approximate)	
Discrete		Continuous	
Simplicial complex	SC_ℓ	Manifold	$\mathcal{M}$
k-cochain	$\boldsymbol{\omega}_k$	k-form	ζ_k
Boundary matrix	$\mathbf{B}_k$	Codifferential operator	δ_k
Coboundary matrix	$\mathbf{B}_k^\top$	Exterior derivative	d_{k-1}
Discrete k-Laplacian	$\mathcal{L}_k$	Laplace-de Rham operator	Δ_k
k-homology space	$\mathcal{H}_k \subseteq \mathbb{R}^{n_k}$	k-homology group	$H_k(\mathcal{M}, \mathbb{R})$

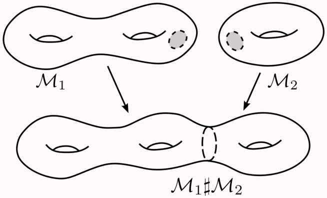
- k-homology space: $\mathcal{H}_k := \ker(\mathcal{L}_k)$ [Lim, 2020, Warner, 2013]
- k^{th} Betti number $\beta_k := \dim(\mathcal{H}_k)$
- k-homology embedding $\mathbf{Y} \in \mathbb{R}^{n_k \times \beta_k}$ is the basis of $\mathcal{H}_k$
- Can estimate a basis of vector fields from $\mathbf{Y}$ for $k = 1$ [Chen et al., 2021]



Connected sum and manifold (prime) decomposition

The connected sum [Lee, 2013] $\mathcal{M} = \mathcal{M}_1 \sharp \mathcal{M}_2$:

1. removing two d-dimensional "disks" from $\mathcal{M}_1$ and $\mathcal{M}_2$ (shaded area)
2. gluing together two manifolds at the boundaries



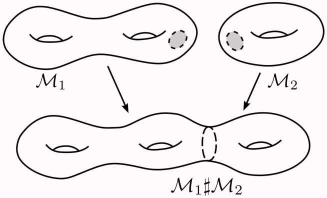
Existence of prime decomposition: factorize a manifold $\mathcal{M} = \mathcal{M}_1 \sharp \cdots \sharp \mathcal{M}_\kappa$ into $\mathcal{M}_i$'s so that $\mathcal{M}_i$ is a prime manifold

- $d = 2$: classification theorem of surfaces [Armstrong, 2013]
- $d = 3$: the uniqueness of the prime decomposition was shown by Kneser-Milnor theorem [Milnor, 1962]
- $d \geqslant 5$: [Bokor et al., 2020] proved the existence of factorization (but they might not be unique)

Connected sum and manifold (prime) decomposition

The connected sum [Lee, 2013] $\mathcal{M} = \mathcal{M}_1 \sharp \mathcal{M}_2$:

1. removing two d-dimensional "disks" from $\mathcal{M}_1$ and $\mathcal{M}_2$ (shaded area)
2. gluing together two manifolds at the boundaries

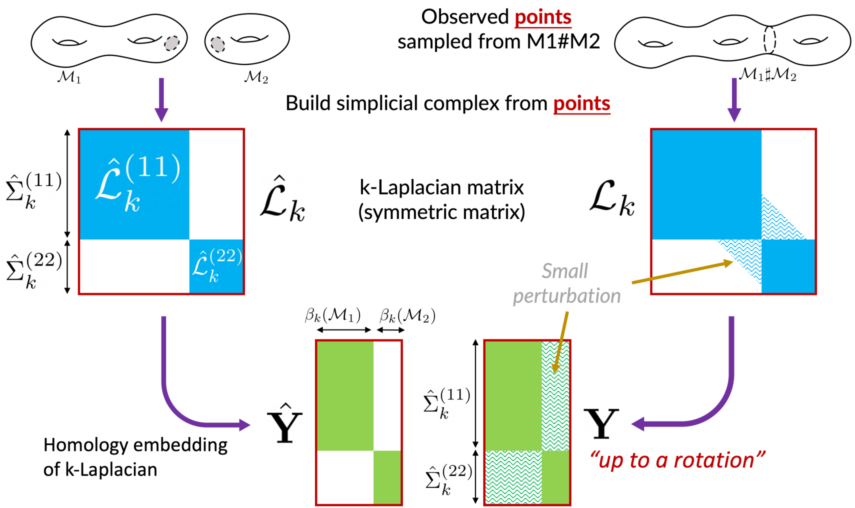


Existence of prime decomposition: factorize a manifold $\mathcal{M} = \mathcal{M}_1 \sharp \cdots \sharp \mathcal{M}_\kappa$ into $\mathcal{M}_i$'s so that $\mathcal{M}_i$ is a prime manifold

- $d = 2$: classification theorem of surfaces [Armstrong, 2013]
- $d = 3$: the uniqueness of the prime decomposition was shown by Kneser-Milnor theorem [Milnor, 1962]
- $d \geqslant 5$: [Bokor et al., 2020] proved the existence of factorization (but they might not be unique)

Problem formulation

NOTATIONS

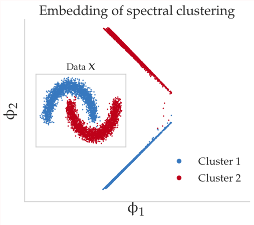


$\mathcal{M}_1$
$\mathcal{M}_2$
Observed points
sampled from M1#M2
$\mathcal{M}_1 \sharp \mathcal{M}_2$
Build simplicial complex from points
$\hat{\Sigma}_k^{(11)}$
$\hat{\Sigma}_k^{(22)}$
$\hat{\mathcal{L}}_k^{(11)}$
$\hat{\mathcal{L}}_k^{(22)}$
$\hat{\mathcal{L}}_k$
k-Laplacian matrix
(symmetric matrix)
$\mathcal{L}_k$
Small
perturbation
$\beta_k(\mathcal{M}_1)$
$\beta_k(\mathcal{M}_2)$
$\hat{\mathbf{Y}}$
$\mathbf{Y}$
Homology embedding
of k-Laplacian
"up to a rotation"

Theoretic and algorithmic aim

Theoretic aim

- Study the geometric properties of $\mathbf{Y}$
  - Recovering the homology basis of each prime manifold $\mathcal{M}_i$
  - Recover $\hat{\mathbf{Y}}$ (localized, support on each $\mathcal{M}_i$) from $\mathbf{Y}$ (coupled, rotation of $\hat{\mathbf{Y}}$)
- Provide an analogous theorem to the OCS [Meilă and Shi, 2001, Ng et al., 2002, Schiebinger et al., 2015] in spectral clustering ($\mathcal{H}_0$)

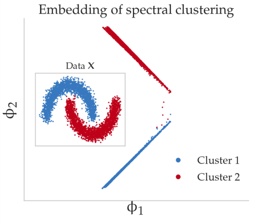


Algorithmic aim

- The null space basis of $\mathcal{L}_k$ is only identifiable up to a unitary matrix
  - $\mathbf{Y}$ is less interpretable than $\mathbf{Z}$!!
- Proposed a data-driven approach to obtain $\mathbf{Z}$ from $\mathbf{Y}$
  - Approximate $\hat{\mathbf{Y}}$ with $\mathbf{Z}$

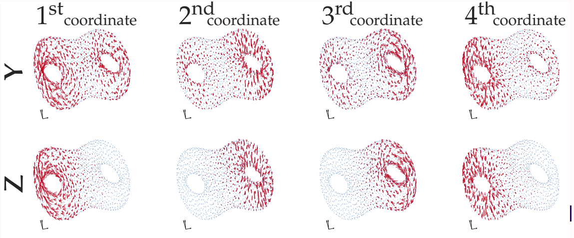
Theoretic aim

- Study the geometric properties of $\mathbf{Y}$
  - Recovering the homology basis of each prime manifold $\mathcal{M}_i$
  - Recover $\hat{\mathbf{Y}}$ (localized, support on each $\mathcal{M}_i$) from $\mathbf{Y}$ (coupled, rotation of $\hat{\mathbf{Y}}$)
- Provide an analogous theorem to the OCS [Meilă and Shi, 2001, Ng et al., 2002, Schiebinger et al., 2015] in spectral clustering ($\mathcal{H}_0$)



Algorithmic aim

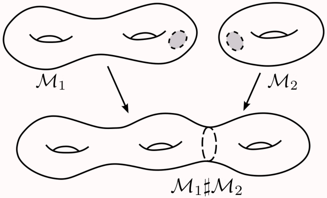
- The null space basis of $\mathcal{L}_k$ is only identifiable up to a unitary matrix
  - $\mathbf{Y}$ is less interpretable than $\mathbf{Z}$!!
- Proposed a data-driven approach to obtain $\mathbf{Z}$ from $\mathbf{Y}$
  - Approximate $\hat{\mathbf{Y}}$ with $\mathbf{Z}$



Connected sum as a matrix perturbation

Assumptions

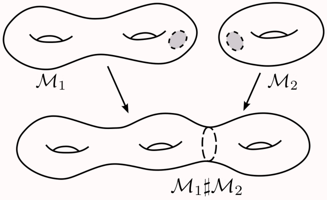
1. Points are sampled from a decomposable manifold
   - κ-fold connected sum: $\mathcal{M} = \mathcal{M}_1 \sharp \cdots \sharp \mathcal{M}_\kappa$
   - $\mathcal{H}_k(\mathrm{SC})$ (discrete) and $H_k(\mathcal{M}, \mathbb{R})$ (continuous) are isomorphic. Also for every $\mathcal{M}_i$
     - Works for any consistent method to build $\mathcal{L}_k$
     - We use our prior work [Chen et al., 2021] for $\mathcal{L}_1$



2. No k-homology class is created/destroyed during the connected sum
   - If $\dim(\mathcal{M}) > k$, then $\mathcal{H}_k(\mathcal{M}_1 \sharp \mathcal{M}_2) \cong \mathcal{H}_k(\mathcal{M}_1) \oplus \mathcal{H}_k(\mathcal{M}_2)$ [Lee, 2013]
   - *[Technical] The eigengap of $\mathcal{L}_k$ is the min of each $\hat{\mathcal{L}}_k^{(ii)}$*: $\delta = \min\{\delta_1, \cdots, \delta_\kappa\}$
3. Sparsely connected manifold
   - Not too many triangles are created/destroyed during connected sum (for $k = 1$)
   - Empirically, the perturbation is small even when $\mathcal{M}$ is not sparsely connected
   - *[Technical] Perturbations of ℓ-simplex set Σ_ℓ are small (ϵ_ℓ and ϵ'_ℓ are small) for $\ell = k, k-1$*

Assumptions

1. Points are sampled from a decomposable manifold
   - κ-fold connected sum: $\mathcal{M} = \mathcal{M}_1 \sharp \cdots \sharp \mathcal{M}_\kappa$
   - $\mathcal{H}_k(\mathrm{SC})$ (discrete) and $H_k(\mathcal{M}, \mathbb{R})$ (continuous) are isomorphic. Also for every $\mathcal{M}_i$
     - Works for any consistent method to build $\mathcal{L}_k$
     - We use our prior work [Chen et al., 2021] for $\mathcal{L}_1$

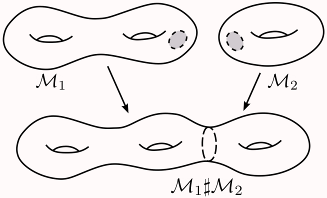


2. No k-homology class is created/destroyed during the connected sum
   - If $\dim(\mathcal{M}) > k$, then $\mathcal{H}_k(\mathcal{M}_1 \sharp \mathcal{M}_2) \cong \mathcal{H}_k(\mathcal{M}_1) \oplus \mathcal{H}_k(\mathcal{M}_2)$ [Lee, 2013]
   - *[Technical] The eigengap of $\mathcal{L}_k$ is the min of each $\hat{\mathcal{L}}_k^{(ii)}$*: $\delta = \min\{\delta_1, \cdots, \delta_\kappa\}$

3. Sparsely connected manifold
   - Not too many triangles are created/destroyed during connected sum (for $k = 1$)
   - Empirically, the perturbation is small even when $\mathcal{M}$ is not sparsely connected
   - *[Technical] Perturbations of ℓ-simplex set Σ_ℓ are small (ϵ_ℓ and ϵ'_ℓ are small) for $\ell = k, k-1$*

Assumptions

1. Points are sampled from a decomposable manifold
   - κ-fold connected sum: $\mathcal{M} = \mathcal{M}_1 \sharp \cdots \sharp \mathcal{M}_\kappa$
   - $\mathcal{H}_k(\mathrm{SC})$ (discrete) and $H_k(\mathcal{M}, \mathbb{R})$ (continuous) are isomorphic. Also for every $\mathcal{M}_i$
     - Works for any consistent method to build $\mathcal{L}_k$
     - We use our prior work [Chen et al., 2021] for $\mathcal{L}_1$



2. No k-homology class is created/destroyed during the connected sum
   - If $\dim(\mathcal{M}) > k$, then $\mathcal{H}_k(\mathcal{M}_1 \sharp \mathcal{M}_2) \cong \mathcal{H}_k(\mathcal{M}_1) \oplus \mathcal{H}_k(\mathcal{M}_2)$ [Lee, 2013]
   - *[Technical] The eigengap of $\mathcal{L}_k$ is the min of each $\hat{\mathcal{L}}_k^{(ii)}$*: $\delta = \min\{\delta_1, \cdots, \delta_\kappa\}$

3. Sparsely connected manifold
   - Not too many triangles are created/destroyed during connected sum (for $k = 1$)
   - Empirically, the perturbation is small even when $\mathcal{M}$ is not sparsely connected
   - *[Technical] Perturbations of ℓ-simplex set Σ_ℓ are small (ϵ_ℓ and ϵ'_ℓ are small) for $\ell = k, k-1$*

THEOREM 1

Under Assumptions 1–3, there exists a unitary matrix $\mathbf{O} \in \mathbb{R}^{\beta_k \times \beta_k}$ *such that*

$$\left\| \mathbf{Y}_{\mathfrak{N}_k,:} - \hat{\mathbf{Y}}_{\mathfrak{N}_k,:}\mathbf{O} \right\|_F^2 \leqslant \frac{8\beta_k \left[\left\| \mathrm{DiffL}_k^{\mathrm{down}} \right\|^2 + \left\| \mathrm{DiffL}_k^{\mathrm{up}} \right\|^2 \right]}{\min\{\delta_1, \cdots, \delta_\kappa\}}, \tag{1}$$

with

$$\left\| \mathrm{DiffL}_k^{\mathrm{down}} \right\|^2 \leqslant \left[2\sqrt{\epsilon_k'} + \epsilon_k' + \left(1 + \sqrt{\epsilon_k'}\right)^2 \sqrt{\epsilon_{k-1}'} + 4\sqrt{\epsilon_{k-1}} \right]^2 (k+1)^2; \text{ and}$$

$$\left\| \mathrm{DiffL}_k^{\mathrm{up}} \right\|^2 \leqslant \left[2\sqrt{\epsilon_k'} + \epsilon_k' + 2\epsilon_k + 4\sqrt{\epsilon_k} \right]^2 (k+2)^2.$$

- Assu. 2: no topology is destroyed/created
- Assu. 3: sparsely connected
- $\mathfrak{N}_k$: bound only simplexes that are **not** altered during connected sum

Subspace perturbation

Theorem 1

Under Assumptions 1–3, there exists a unitary matrix $\mathbf{O} \in \mathbb{R}^{\beta_k \times \beta_k}$ *such that*

$$\left\| \mathbf{Y}_{\mathfrak{N}_k,:} - \hat{\mathbf{Y}}_{\mathfrak{N}_k,:}\mathbf{O} \right\|_F^2 \leqslant \frac{8\beta_k \left[\left\| \mathrm{DiffL}_k^{\mathrm{down}} \right\|^2 + \left\| \mathrm{DiffL}_k^{\mathrm{up}} \right\|^2 \right]}{\min\{\delta_1, \cdots, \delta_\kappa\}}, \tag{1}$$

with

$$\left\| \mathrm{DiffL}_k^{\mathrm{down}} \right\|^2 \leqslant \left[2\sqrt{\epsilon_k'} + \epsilon_k' + \left(1 + \sqrt{\epsilon_k'}\right)^2 \sqrt{\epsilon_{k-1}'} + 4\sqrt{\epsilon_{k-1}} \right]^2 (k+1)^2; \text{ and}$$

$$\left\| \mathrm{DiffL}_k^{\mathrm{up}} \right\|^2 \leqslant \left[2\sqrt{\epsilon_k'} + \epsilon_k' + 2\epsilon_k + 4\sqrt{\epsilon_k} \right]^2 (k+2)^2.$$

- Assu. 2: no topology is destroyed/created
- Assu. 3: sparsely connected
- $\mathfrak{N}_k$: bound only simplexes that are **not** altered during connected sum

Subspace perturbation

Theorem 1

Under Assumptions 1–3, there exists a unitary matrix $\mathbf{O} \in \mathbb{R}^{\beta_k \times \beta_k}$ *such that*

$$\left\|\mathbf{Y}_{\mathfrak{N}_k,:} - \hat{\mathbf{Y}}_{\mathfrak{N}_k,:}\mathbf{O}\right\|_F^2 \leqslant \frac{8\beta_k \left[\left\|\mathrm{DiffL}_k^{\mathrm{down}}\right\|^2 + \left\|\mathrm{DiffL}_k^{\mathrm{up}}\right\|^2\right]}{\min\{\delta_1, \cdots, \delta_K\}}, \quad (1)$$

with

$$\left\|\mathrm{DiffL}_k^{\mathrm{down}}\right\|^2 \leqslant \left[2\sqrt{\epsilon_k'} + \epsilon_k' + \left(1 + \sqrt{\epsilon_k'}\right)^2 \sqrt{\epsilon_{k-1}'} + 4\sqrt{\epsilon_{k-1}}\right]^2 (k+1)^2; \text{ and}$$

$$\left\|\mathrm{DiffL}_k^{\mathrm{up}}\right\|^2 \leqslant \left[2\sqrt{\epsilon_k'} + \epsilon_k' + 2\epsilon_k + 4\sqrt{\epsilon_k}\right]^2 (k+2)^2.$$

- Assu. 2: no topology is destroyed/created
- Assu. 3: sparsely connected
- $\mathfrak{N}_k$: bound only simplexes that are **not** altered during connected sum

Subspace perturbation

Theorem 1

Under Assumptions 1–3, there exists a unitary matrix $\mathbf{O} \in \mathbb{R}^{\beta_k \times \beta_k}$ *such that*

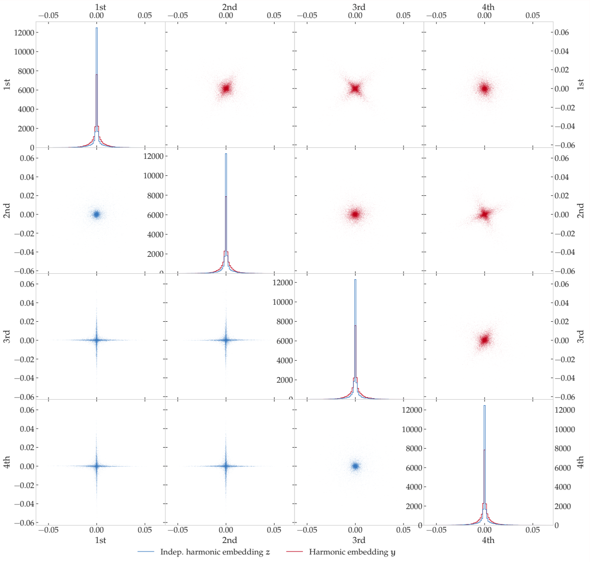
$$\left\| \mathbf{Y}_{\mathfrak{N}_k,:} - \hat{\mathbf{Y}}_{\mathfrak{N}_k,:}\mathbf{O} \right\|_F^2 \leqslant \frac{8\beta_k \left[\left\| \mathrm{DiffL}_k^{\mathrm{down}} \right\|^2 + \left\| \mathrm{DiffL}_k^{\mathrm{up}} \right\|^2 \right]}{\min\{\delta_1, \cdots, \delta_\kappa\}}, \tag{1}$$

with

$$\left\| \mathrm{DiffL}_k^{\mathrm{down}} \right\|^2 \leqslant \left[2\sqrt{\epsilon_k'} + \epsilon_k' + \left(1 + \sqrt{\epsilon_k'}\right)^2 \sqrt{\epsilon_{k-1}'} + 4\sqrt{\epsilon_{k-1}} \right]^2 (k+1)^2; \text{ and}$$

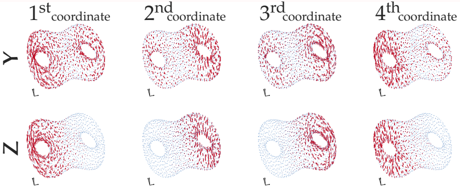
$$\left\| \mathrm{DiffL}_k^{\mathrm{up}} \right\|^2 \leqslant \left[2\sqrt{\epsilon_k'} + \epsilon_k' + 2\epsilon_k + 4\sqrt{\epsilon_k} \right]^2 (k+2)^2.$$

- **Assu. 2:** no topology is destroyed/created
- **Assu. 3:** sparsely connected
- $\mathfrak{N}_k$: bound only simplexes that are **not** altered during connected sum



Input: $\mathbf{Y}$ (coupled)

Output: $\mathbf{Z}$ (localized, approx. of $\hat{\mathbf{Y}}$)



$$\left\| \mathbf{Y}_{\mathfrak{N}_k,:} - \hat{\mathbf{Y}}_{\mathfrak{N}_k,:} \mathbf{O} \right\|_F^2 \leqslant \frac{8\beta_k \cdot (\cdots)}{\min\{\delta_1, \cdots, \delta_K\}}$$

Estimate $\mathbf{O}$ with Independent Component Analysis (ICA)

Applications

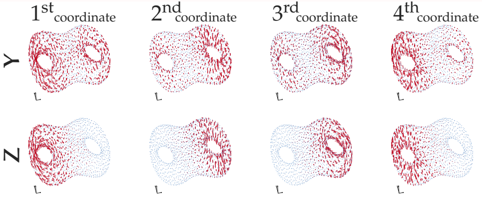
Applications

- Classifying any 2-manifold
  - $\mathbb{S}^1 \sharp \mathbb{S}^1 \neq \mathbb{T}^2$ even though $\beta_1 = 2$ for both
  - **Proposition 4**: 1-homology embedding of $\mathbb{T}^m$ is an m-dimensional ellipsoid

- Visualize the basis of harmonic vector fields
- Higher-order simplex clustering [Ebli and Spreemann, 2019]
  - Theorem 1 supports their use of subspace clustering algorithm
- Shortest homologous loop detection
  - **Proposition 3**: a non-trivial loop corresponding to the i^{th} column of the homology embedding can be obtained using Dijkstra algorithm
  - Using the factorized homology embedding $\mathbf{Z}$ ensures that each loop corresponds to a single homology class

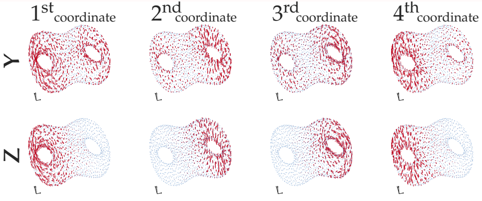
Applications

- Classifying any 2-manifold
  - $\mathbb{S}^1 \sharp \mathbb{S}^1 \neq \mathbb{T}^2$ even though $\beta_1 = 2$ for both
  - **Proposition 4**: 1-homology embedding of $\mathbb{T}^m$ is an m-dimensional ellipsoid
- Visualize the basis of harmonic vector fields
- Higher-order simplex clustering [Ebli and Spreemann, 2019]
  - Theorem 1 supports their use of subspace clustering algorithm
- Shortest homologous loop detection
  - **Proposition 3**: a non-trivial loop corresponding to the i^{th} column of the homology embedding can be obtained using Dijkstra algorithm
  - Using the factorized homology embedding $\mathbf{Z}$ ensures that each loop corresponds to a single homology class



Applications

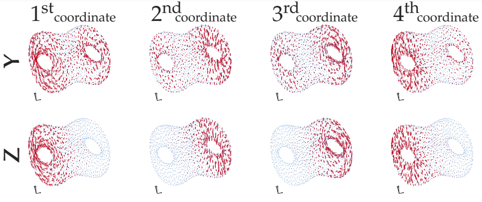
- Classifying any 2-manifold
  - $\mathbb{S}^1 \sharp \mathbb{S}^1 \neq \mathbb{T}^2$ even though $\beta_1 = 2$ for both
  - **Proposition 4**: 1-homology embedding of $\mathbb{T}^m$ is an m-dimensional ellipsoid
- Visualize the basis of harmonic vector fields
- Higher-order simplex clustering [Ebli and Spreemann, 2019]
  - Theorem 1 supports their use of subspace clustering algorithm



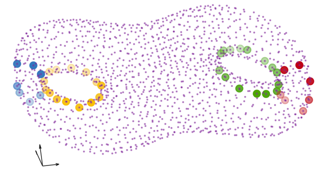
- Shortest homologous loop detection
  - **Proposition 3**: a non-trivial loop corresponding to the i^{th} column of the homology embedding can be obtained using Dijkstra algorithm
  - Using the factorized homology embedding **Z** ensures that each loop corresponds to a single homology class

Applications

- Classifying any 2-manifold
  - $\mathbb{S}^1 \sharp \mathbb{S}^1 \neq \mathbb{T}^2$ even though $\beta_1 = 2$ for both
  - **Proposition 4**: 1-homology embedding of $\mathbb{T}^m$ is an m-dimensional ellipsoid
- Visualize the basis of harmonic vector fields
- Higher-order simplex clustering [Ebli and Spreemann, 2019]
  - Theorem 1 supports their use of subspace clustering algorithm



- Shortest homologous loop detection
  - **Proposition 3**: a non-trivial loop corresponding to the i^{th} column of the homology embedding can be obtained using Dijkstra algorithm
  - Using the factorized homology embedding $\mathbf{Z}$ ensures that each loop corresponds to a single homology class

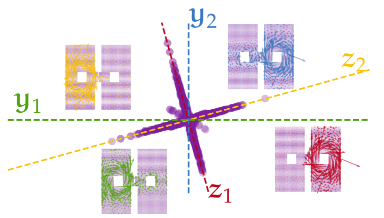


Experiments

Synthetic Manifolds: Two Disjoint Holes $\mathbb{S}^1 \sharp \mathbb{S}^1$ and Tori $\mathbb{T}^m$

Two disjoint holes $\mathbb{S}^1 \sharp \mathbb{S}^1$:

- Inset: estimated vector field from the corresponding basis with [Chen et al., 2021]
- Red and yellow (z_1 and z_2) are more localized than green and blue

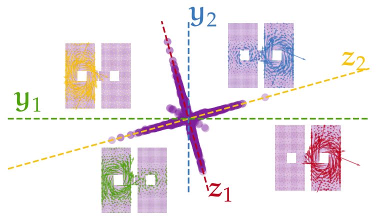


m-tori $\mathbb{T}^m$:

- Homology embedding of $\mathbb{T}^2$ is different from that of $\mathbb{S}^1 \sharp \mathbb{S}^1$
  - Classify them by Proposition 4
- $\mathbf{Z}$ of $\mathbb{T}^3$ is an ellipsoid

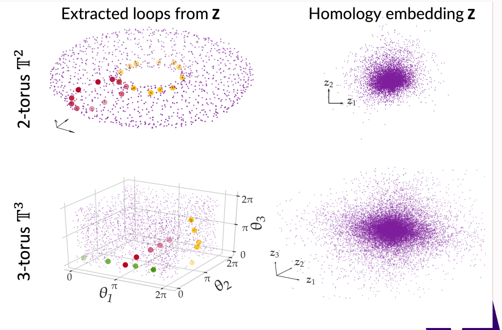
Two disjoint holes $\mathbb{S}^1 \sharp \mathbb{S}^1$:

- Inset: estimated vector field from the corresponding basis with [Chen et al., 2021]
- Red and yellow (z_1 and z_2) are more localized than green and blue

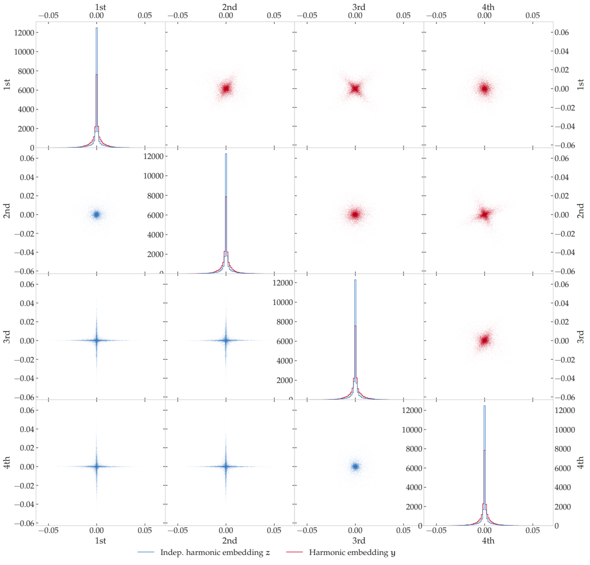


m-tori $\mathbb{T}^m$:

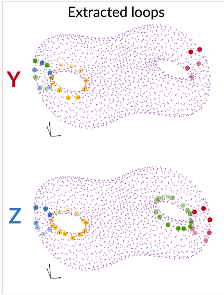
- Homology embedding of $\mathbb{T}^2$ is different from that of $\mathbb{S}^1 \sharp \mathbb{S}^1$
  - Classify them by Proposition 4
- $\mathbf{Z}$ of $\mathbb{T}^3$ is an ellipsoid



SYNTHETIC MANIFOLDS: COMPLEX SURFACES

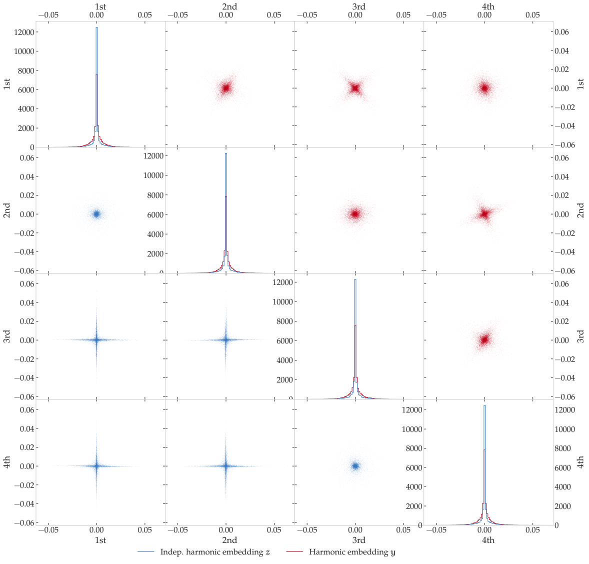


Genus-2 surface:

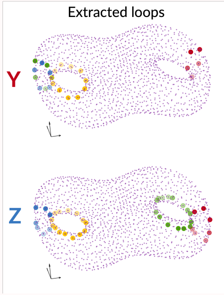


Concatenation of 4 tori:

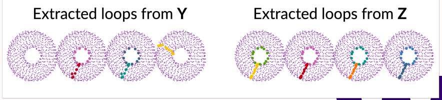
Synthetic manifolds: complex surfaces



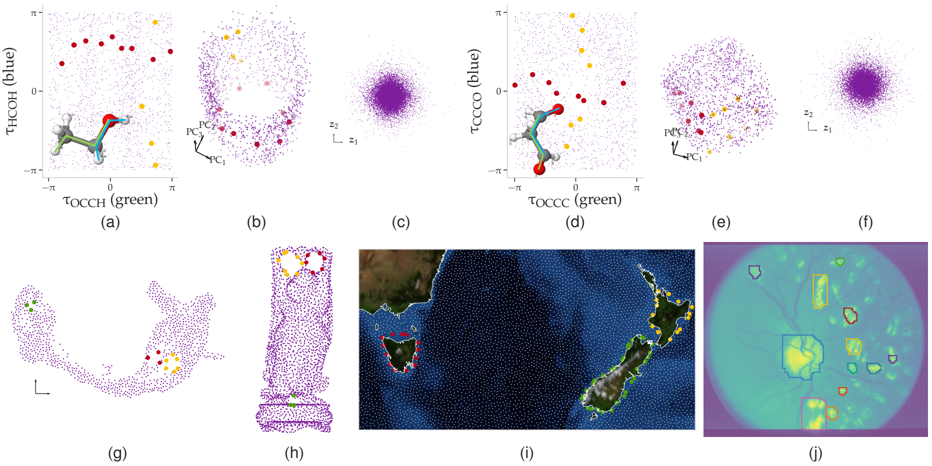
Genus-2 surface:



Concatenation of 4 tori:

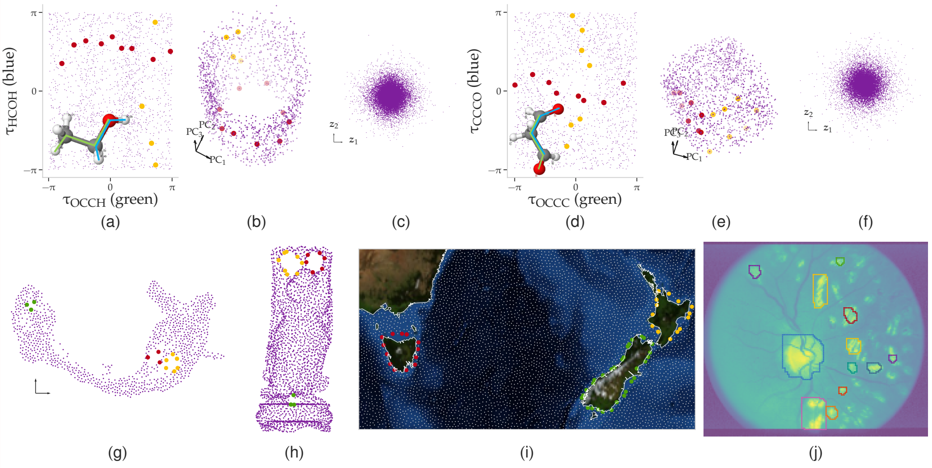


Real datasets



- Fig. (j): our framework can be extended to images with cubical complex

Real datasets



- Fig. (j): our framework can be extended to images with cubical complex

Discussion

Prior work

Geometry/shape for the $\mathcal{H}_0$ embedding.

- Pivotal for spectral clustering and inference algorithms for the stochastic block models
  - Using matrix perturbation theory [Ng et al., 2002, Wan and Meila, 2015, von Luxburg, 2007]
  - Under the assumption of a mixture model [Schiebinger et al., 2015]

Higher-order homology embeddings ($k > 0$).

- Reported empirically that the homology embedding is approximately distributed on the union (directed sum) of subspaces [Ebli and Spreemann, 2019]
  - Subspace clustering algorithms [Kailing et al., 2004] were applied to cluster edges/triangles

Contributions

- Generalize the study of embedding of the spectral clustering to higher-order homology embedding of $\mathcal{H}_k$
- Our analysis is made possible by expressing the κ-fold connected sum as a matrix perturbation
  - Theoretical: the k-homology embedding can be approximately factorized into parts, with each corresponding to a prime manifold given a small perturbation
  - Algorithmic: identify each decoupled subspace using ICA
  - Easy to extend to cubical complexes in image analysis
- Applications in shortest homologous loop detection, classifying any 2-dimensional manifold, and visualizing harmonic vector fields.
- Support our theoretical claims by comprehensive experiments on synthetic and real datasets

Future work[1]

1. Extend our framework to a multiple spatial resolution approach
   - The persistent spectral methods [Wang et al., 2020, Meng and Xia, 2021]

2. Explore the connection between the proposed framework and the disentangled representations [Zhou et al., 2020]

3. Investigate the success/failure conditions of the proposed spectral homologous loop detection algorithm

[1] We thank the anonymous reviewers for suggesting some of these directions to explore.

Thank you very much!

References

[Armstrong, 2013] Armstrong, M. A. (2013).
***Basic Topology*.**
Springer Science & Business Media.

[Berry and Sauer, 2019] Berry, T. and Sauer, T. (2019).
Consistent manifold representation for topological data analysis.
Foundations of Data Science, 1(1):1.

[Bokor et al., 2020] Bokor, I., Crowley, D., Friedl, S., Hebestreit, F., Kasprowski, D., Land, M., and Nicholson, J. (2020).
Connected sum decompositions of high-dimensional manifolds.
arXiv:1909.02628 [math].

[Chen et al., 2021] Chen, Y.-C., Meilă, M., and Kevrekidis, I. G. (2021).
Helmholtzian Eigenmap: Topological feature discovery & edge flow learning from point cloud data.
arXiv:2103.07626 [stat.ML].

[Ebli and Spreemann, 2019] Ebli, S. and Spreemann, G. (2019).
A Notion of Harmonic Clustering in Simplicial Complexes.
2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), pages 1083–1090.

[Eckmann, 1944] Eckmann, B. (1944).
Harmonische Funktionen und Randwertaufgaben in einem Komplex.
Commentarii mathematici Helvetici, 17:240–255.

[Horak and Jost, 2013] Horak, D. and Jost, J. (2013).
Spectra of combinatorial Laplace operators on simplicial complexes.
Advances in Mathematics, 244:303–336.

[Kailing et al., 2004] Kailing, K., Kriegel, H.-P., and Kröger, P. (2004).
Density-connected subspace clustering for high-dimensional data.
In *Proceedings of the 2004 SIAM International Conference on Data Mining*, pages 246–256. SIAM.

References III

[Lee, 2013] Lee, J. M. (2013).
Introduction to Smooth Manifolds.
In *Introduction to Smooth Manifolds*, pages 1–31. Springer.

[Lim, 2020] Lim, L.-H. (2020).
Hodge laplacians on graphs.
Siam Review, 62(3):685–715.

[Meilă and Shi, 2001] Meilă, M. and Shi, J. (2001).
A random walks view of spectral segmentation.
In *International Workshop on Artificial Intelligence and Statistics*, pages 203–208. PMLR.

[Meng and Xia, 2021] Meng, Z. and Xia, K. (2021).
Persistent spectral–based machine learning (PerSpect ML) for protein-ligand binding affinity prediction.
Science Advances, 7(19):eabc5329.

References IV

[Milnor, 1962] Milnor, J. (1962).
A unique decomposition theorem for 3-manifolds.
American Journal of Mathematics, 84(1):1–7.

[Ng et al., 2002] Ng, A. Y., Jordan, M. I., and Weiss, Y. (2002).
On spectral clustering: Analysis and an algorithm.
Advances in neural information processing systems, 2:849–856.

[Schaub et al., 2020] Schaub, M. T., Benson, A. R., Horn, P., Lippner, G., and Jadbabaie, A. (2020).
Random walks on simplicial complexes and the normalized Hodge 1-Laplacian.
SIAM Review, 62(2):353–391.

[Schiebinger et al., 2015] Schiebinger, G., Wainwright, M. J., and Yu, B. (2015).
The geometry of kernelized spectral clustering.
Annals of Statistics, 43(2):819–846.

References V

[von Luxburg, 2007] von Luxburg, U. (2007).
A Tutorial on Spectral Clustering.
arXiv:0711.0189 [cs].

[Wan and Meila, 2015] Wan, Y. and Meila, M. (2015).
A class of network models recoverable by spectral clustering.
In *NIPS*, pages 3285–3293.

[Wang et al., 2020] Wang, R., Nguyen, D. D., and Wei, G.-W. (2020).
Persistent spectral graph.
International journal for numerical methods in biomedical engineering, 36(9):e3376.

[Warner, 2013] Warner, F. W. (2013).
***Foundations of Differentiable Manifolds and Lie Groups*, volume 94.**
Springer Science & Business Media.

References VI

[Yu et al., 2015] Yu, Y., Wang, T., and Samworth, R. J. (2015).
A useful variant of the Davis–Kahan theorem for statisticians.
Biometrika, 102(2):315–323.

[Zhou et al., 2020] Zhou, S., Zelikman, E., Lu, F., Ng, A. Y., Carlsson, G., and Ermon, S. (2020).
Evaluating the disentanglement of deep generative models through manifold topology.
arXiv preprint arXiv:2006.03680.

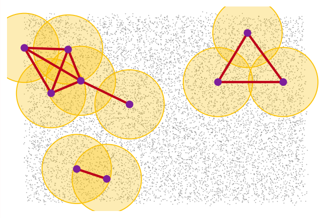
Backup slides

Backup slides

Simplicial complexes, cochains, and boundary matrices

High-dimensional i.i.d. samples and neighborhood graph

- Observed data $x_i \in \mathbb{R}^D$ for $i = 1, \cdots, n$ sampled (i.i.d.) from a d-manifold
  - Called a point cloud
- Local low dimensional geometry is encoded in local distances, triangles, tetrahedra, etc.
  - Represented by a neighborhood graph



δ-radius neighborhood graph

$G = (V, E)$ with

- the vertex set V on every x_i's (index set)
- the edge set E being

$$E = \{(i, j) \in V^2 : \|x_i - x_j\|_2 \leqslant \delta\}.$$

Simplicial and cubical complexes — I

Simplicial complex SC

An SC is a set of simplices so that:

1. Every face of a simplex from SC is also in SC
2. $\sigma_1 \cap \sigma_2$ for any $\sigma_1, \sigma_2 \in$ SC is a face of both σ_1 and σ_2

- Σ_ℓ is the collection of ℓ-simplices σ_ℓ, then

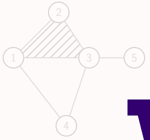
$$SC_k = (\Sigma_\ell)_{\ell=0}^{k} = (\Sigma_0, \Sigma_1, \cdots, \Sigma_k)$$

- The cardinality of Σ_ℓ is $n_\ell = |\Sigma_\ell|$

Not an SC

Remark.

1. A graph is: $G = SC_1 = (V, E) = (\Sigma_0, \Sigma_1)$
2. We *mostly* focus on $SC_2 = (V, E, T) = (\Sigma_0, \Sigma_1, \Sigma_2)$



An SC

Simplicial and cubical complexes — I

Simplicial complex SC

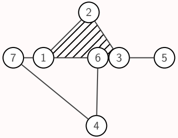
An **SC** is a set of simplices so that:

1. Every face of a simplex from **SC** is also in **SC**
2. $\sigma_1 \cap \sigma_2$ for any $\sigma_1, \sigma_2 \in$ **SC** is a face of both σ_1 and σ_2

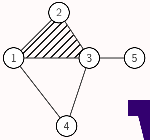
- Σ_ℓ is the collection of ℓ-simplices σ_ℓ, then

$$SC_k = (\Sigma_\ell)_{\ell=0}^{k} = (\Sigma_0, \Sigma_1, \cdots, \Sigma_k)$$

- The cardinality of Σ_ℓ is $n_\ell = |\Sigma_\ell|$



Not an **SC**



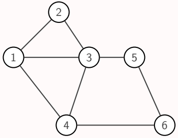
An **SC**

Remark.

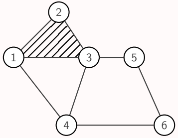
1. A graph is: $G = SC_1 = (V, E) = (\Sigma_0, \Sigma_1)$
2. We *mostly* focus on $SC_2 = (V, E, T) = (\Sigma_0, \Sigma_1, \Sigma_2)$

CLIQUE COMPLEX

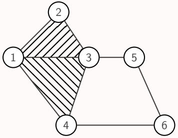
A clique complex of a graph $\mathbf{G} = (V, E)$ is a simplicial complex $\mathbf{SC}_k = (\Sigma_0, \cdots, \Sigma_k)$, with the ℓ-th simplex set Σ_ℓ being the set of all ℓ-cliques



Graph $\mathbf{G}$



NOT a clique complex of $\mathbf{G}$



A clique complex of $\mathbf{G}$

Remark. The clique complex built from δ-radius graph $:=$ Vietoris-Rips (VR) complex

CUBICAL COMPLEX *(INFORMAL)*

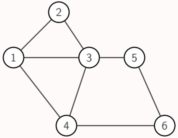
A cubical complex $\mathbf{CB}_k = (K_0, \cdots, K_k)$ is a collection of sets K_ℓ of ℓ-cubes

Remark. $\mathbf{CB}_k$ is widely used for image datasets

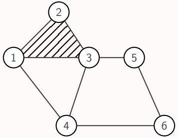
Simplicial and cubical complexes — II

Clique complex

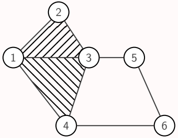
A clique complex of a graph $G = (V, E)$ is a simplicial complex $SC_k = (\Sigma_0, \cdots, \Sigma_k)$, with the ℓ-th simplex set Σ_ℓ being the set of all ℓ-cliques



Graph G



NOT a clique complex of G



A clique complex of G

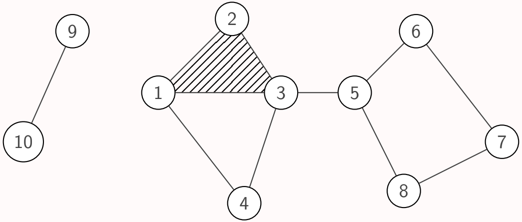
Remark. The clique complex built from δ-radius graph $:=$ Vietoris-Rips (VR) complex

Cubical complex *(informal)*

A cubical complex $CB_k = (K_0, \cdots, K_k)$ is a collection of sets K_ℓ of ℓ-cubes

Remark. CB_k is widely used for image datasets

An SC_2 defines (co-)boundary matrices $\mathbf{B}_1$ and $\mathbf{B}_2$



$\mathbf{B}_1 =$

# of col = n_1; # of row = n_0

	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(3, 4)	(3, 5)	(5, 6)	(5, 8)	(6, 7)	(7, 8)	(9, 10)
1	1	1	1	0	0	0	0	0	0	0	0
2	-1	0	0	1	0	0	0	0	0	0	0
3	0	-1	0	-1	1	1	0	0	0	0	0
4	0	0	-1	0	-1	0	0	0	0	0	0
5	0	0	0	0	0	-1	1	1	0	0	0
6	0	0	0	0	0	0	-1	0	1	0	0
7	0	0	0	0	0	0	0	0	-1	1	0
8	0	0	0	0	0	0	0	-1	0	-1	0
9	0	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	0	0	-1

$\mathbf{B}_2 =$

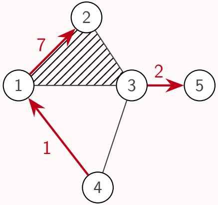
# of col = n_2; # of row = n_1

	(1, 2, 3)
(1, 2)	1
(1, 3)	-1
(1, 4)	0
(2, 3)	1
(3, 4)	0
(3, 5)	0
(5, 6)	0
(5, 8)	0
(6, 7)	0
(7, 8)	0

k-COCHAIN

An edge flow (1-cochain) $\boldsymbol{\omega}_1$ is a flow on edges (1-simplex) of SC/CB

- $\boldsymbol{\omega}_1 = \sum_i \omega_{1,i} e_i$, where $e_i \in E$
- Can further denote by $\boldsymbol{\omega}_1 = (\omega_{1,1}, \cdots, \omega_{1,n_1})^\top \in \mathbb{R}^{n_1}$
  - Set of $\pm$ weights on edges
- Space of $\boldsymbol{\omega}$ ($:= \mathcal{C}_1$) is isomorphic to $\mathbb{R}^{n_1}$



Example. $\omega_1 = 7 \cdot [1,2] + 2 \cdot [3,5] + (-1) \cdot [1,4]$

$$\omega_1 = \begin{bmatrix} 7 & 0 & -1 & 0 & 0 & 2 \\ [1,2] & [1,3] & [1,4] & [2,3] & [3,4] & [3,5] \end{bmatrix} \in \mathbb{R}^6$$

$\omega_k :=$ HIGHER-ORDER GENERALIZATION OF ω_1

A k-cochain ω_k is a flow on k-simplex of SC/CB

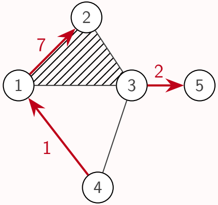
k-cochain

An edge flow (1-cochain) $\boldsymbol{\omega}_1$ is a flow on edges (1-simplex) of SC/CB

- $\boldsymbol{\omega}_1 = \sum_i \omega_{1,i} e_i$, where $e_i \in E$
- Can further denote by $\boldsymbol{\omega}_1 = (\omega_{1,1}, \cdots, \omega_{1,n_1})^\top \in \mathbb{R}^{n_1}$
  - Set of $\pm$ weights on edges
- Space of $\boldsymbol{\omega}$ ($:= \mathcal{C}_1$) is isomorphic to $\mathbb{R}^{n_1}$

Example. $\boldsymbol{\omega}_1 = 7 \cdot [1,2] + 2 \cdot [3,5] + (-1) \cdot [1,4]$

$$\boldsymbol{\omega}_1 = \begin{bmatrix} 7 & 0 & -1 & 0 & 0 & 2 \\ [1,2] & [1,3] & [1,4] & [2,3] & [3,4] & [3,5] \end{bmatrix} \in \mathbb{R}^6$$



$\omega_k :=$ Higher-order generalization of ω_1

A k-cochain $\boldsymbol{\omega}_k$ is a flow on k-simplex of SC/CB

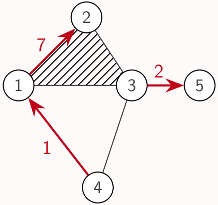
k-COCHAIN

An edge flow (1-cochain) $\boldsymbol{\omega}_1$ is a flow on edges (1-simplex) of SC/CB

- $\boldsymbol{\omega}_1 = \sum_i \omega_{1,i} e_i$, where $e_i \in E$
- Can further denote by $\boldsymbol{\omega}_1 = (\omega_{1,1}, \cdots, \omega_{1,n_1})^\top \in \mathbb{R}^{n_1}$
  - Set of $\pm$ weights on edges
- Space of $\boldsymbol{\omega}$ ($:= \mathcal{C}_1$) is isomorphic to $\mathbb{R}^{n_1}$

Example. $\omega_1 = 7 \cdot [1,2] + 2 \cdot [3,5] + (-1) \cdot [1,4]$

$$\omega_1 = \begin{bmatrix} 7 & 0 & -1 & 0 & 0 & 2 \\ [1,2] & [1,3] & [1,4] & [2,3] & [3,4] & [3,5] \end{bmatrix} \in \mathbb{R}^6$$



$\boldsymbol{\omega}_k :=$ Higher-order generalization of $\boldsymbol{\omega}_1$

A k-cochain $\boldsymbol{\omega}_k$ is a flow on k-simplex of SC/CB

Backup slides

The discrete k-Laplacian

Higher-order Laplacians

k-Laplacians

Unnormalized k*-Laplacian* [Eckmann, 1944]:

$$\mathbf{L}_k = \underbrace{\mathbf{B}_k^\top \mathbf{B}_k}_{\mathbf{L}_k^{\text{down}}} + \underbrace{\mathbf{B}_{k+1}\mathbf{B}_{k+1}^\top}_{\mathbf{L}_k^{\text{up}}};$$

Random-walk k*-Laplacian* [Horak and Jost, 2013]:

$$\mathcal{L}_k = \underbrace{\mathbf{B}_k^\top \mathbf{W}_{k-1}^{-1}\mathbf{B}_k \mathbf{W}_k}_{\mathcal{L}_k^{\text{down}}} + \underbrace{\mathbf{W}_1^{-1}\mathbf{B}_2\mathbf{W}_2\mathbf{B}_2^\top}_{\mathcal{L}_k^{\text{up}}};$$

Symmetrized k*-Laplacian* [Schaub et al., 2020]:

$$\mathcal{L}_k^s = \underbrace{\mathbf{A}_k^\top \mathbf{A}_k}_{\mathcal{L}_k^{s,\text{down}}} + \underbrace{\mathbf{A}_{k+1}\mathbf{A}_{k+1}^\top}_{\mathcal{L}_k^{s,\text{up}}}.$$

- $\mathbf{A}_\ell \coloneqq \mathbf{W}_{\ell-1}^{-1/2}\mathbf{B}_\ell \mathbf{W}_\ell^{1/2}$ (for $\ell = k, k+1$) is the normalized boundary matrix
- $\mathcal{L}_k^s = \mathbf{W}_k^{1/2}\mathcal{L}_k \mathbf{W}_k^{-1/2}$ has the same spectrum as $\mathcal{L}_k$ [Schaub et al., 2020]

The *up*- and *down*-1-Laplacian

$$\mathcal{L}_1^{\text{down}} = \mathbf{B}_1^\top \, W_0^{-1} \, \mathbf{B}_1 \, W_1 \in \mathbb{R}^{n_1 \times n_1}$$

$n_1 \times n_0$ $\quad n_0 \times n_0$ $\quad n_0 \times n_1$ $\quad n_1 \times n_1$

$$\mathcal{L}_1^{\text{up}} = W_1^{-1} \, \mathbf{B}_2 \, W_2 \, \mathbf{B}_2^\top \in \mathbb{R}^{n_1 \times n_1}$$

$n_1 \times n_1$ $\quad n_1 \times n_2$ $\quad n_2 \times n_2$ $\quad n_2 \times n_1$

k-Laplacians are the extensions of graph Laplacians

- $\mathbf{L}_0 = \mathbf{B}_1\mathbf{B}_1^\top$ is the unnormalized graph Laplacian:

$$\mathbf{L}_0 = \mathbf{B}_1\mathbf{B}_1^\top = \begin{cases} \deg(i) & \text{if } i = j \\ -1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases} = \mathbf{D} - \mathbf{A}$$

By letting $\mathbf{W}_0 = \text{diag}(|\mathbf{B}_1|\mathbf{W}_1\mathbf{1}) = \text{diag}\left(\left[\sum_j w_{ij}\right]_{i=1}^n\right) = \mathbf{D}\ldots$

- $\mathcal{L}_0 = \mathbf{W}_0^{-1}\mathbf{B}_1\mathbf{W}_1\mathbf{B}_1^\top$ is the random-walk graph Laplacian:

$$\mathcal{L}_0 = \mathbf{D}^{-1}\mathbf{B}_1\mathbf{W}_1\mathbf{B}_1^\top = \begin{cases} 1 & \text{if } i = j \\ -\frac{1}{\deg(i)} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}$$

k-Laplacians are the extensions of graph Laplacians

- $\mathbf{L}_0 = \mathbf{B}_1\mathbf{B}_1^\top$ is the unnormalized graph Laplacian:

$$\mathbf{L}_0 = \mathbf{B}_1\mathbf{B}_1^\top = \begin{cases} \deg(i) & \text{if } i = j \\ -1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases} = \mathbf{D} - \mathbf{A}$$

By letting $\mathbf{W}_0 = \mathrm{diag}(|\mathbf{B}_1|\mathbf{W}_1\mathbf{1}) = \mathrm{diag}\left(\left[\sum_j w_{ij}\right]_{i=1}^n\right) = \mathbf{D}$...

- $\mathcal{L}_0 = \mathbf{W}_0^{-1}\mathbf{B}_1\mathbf{W}_1\mathbf{B}_1^\top$ is the random-walk graph Laplacian:

$$\mathcal{L}_0 = \mathbf{D}^{-1}\mathbf{B}_1\mathbf{W}_1\mathbf{B}_1^\top = \begin{cases} 1 & \text{if } i = j \\ -\frac{1}{\deg(i)} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}$$

Backup slides

Hodge Laplacian, differential geometry, and topology

k-HOMOLOGY SPACE

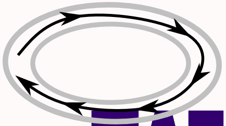
Harmonic vector space

The harmonic vector space $\mathcal{H}_k \subseteq \mathbb{R}^{n_1}$ is a subspace of the k-cochain defined as the null of $\mathcal{L}_k$

$$\mathcal{H}_k \coloneqq \{\omega \in \mathbb{R}^{n_k} : \mathcal{L}_k \omega = 0\}.$$

Remark. Similar definition works for $\mathbf{L}_k$ or $\mathcal{L}_k^s$, as well as its continuous counterpart (using k-differential forms and Δ_k)

- The k-th homology space $H_k \coloneqq \ker(\mathbf{B}_k)/\operatorname{im}(\mathbf{B}_{k+1})$
- $\mathcal{H}_k \cong H_k$ [Lim, 2020, Warner, 2013]
- The k-th Betti number $\beta_k \coloneqq \dim(\mathcal{H}_k) = \dim(\ker(\mathcal{L}_k))$



k-HOMOLOGY SPACE

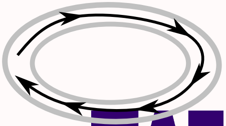
HARMONIC VECTOR SPACE

The harmonic vector space $\mathcal{H}_k \subseteq \mathbb{R}^{n_1}$ is a subspace of the k-cochain defined as the null of $\boldsymbol{\mathcal{L}}_k$

$$\mathcal{H}_k \coloneqq \{\boldsymbol{\omega} \in \mathbb{R}^{n_k} : \boldsymbol{\mathcal{L}}_k \boldsymbol{\omega} = 0\}.$$

Remark. Similar definition works for $\mathbf{L}_k$ or $\boldsymbol{\mathcal{L}}_k^s$, as well as its continuous counterpart (using k-differential forms and Δ_k)

- The k-th homology space $\mathrm{H}_k \coloneqq \ker(\mathbf{B}_k)/\mathrm{im}(\mathbf{B}_{k+1})$
- $\mathcal{H}_k \cong \mathrm{H}_k$ [Lim, 2020, Warner, 2013]
- The k-th Betti number $\beta_k \coloneqq \dim(\mathcal{H}_k) = \dim(\ker(\boldsymbol{\mathcal{L}}_k))$



Connection to the continuous operators

(Finite samples from $\mathcal{M}$)		(Want to approximate)	
Discrete		Continuous	
Simplicial/Cubical complex	SC_ℓ (or CB_ℓ)	Manifold	$\mathcal{M}$
k-cochain	$\boldsymbol{\omega}_k$	k-form	ζ_k
Boundary matrix	$\mathbf{B}_k$	Codifferential operator	δ_k
Coboundary matrix	$\mathbf{B}_k^\top$	Exterior derivative	d_{k-1}
Discrete k-Laplacian	$\boldsymbol{\mathcal{L}}_k$	Laplace-de Rham operator	Δ_k
k-homology space	$\mathcal{H}_k \subseteq \mathbb{R}^{n_k}$	k-homology group	$H_k(\mathcal{M}, \mathbb{R})$

Backup slides

Boundary matrices

Boundary matrix

A boundary matrix $\mathbf{B}_k \in \mathbb{R}^{n_k \times n_{k-1}}$ maps a k-simplex to its $(k-1)$-th faces

- With $[x, y, z] \in T$, $\mathbf{B}_1$ and $\mathbf{B}_2$ are defined as:

$$[\mathbf{B}_1]_{a,xy} = \begin{cases} 1 & \text{if } a = x \\ -1 & \text{if } a = y \\ 0 & \text{otherwise} \end{cases} \quad ; \quad [\mathbf{B}_2]_{ab,xyz} = \begin{cases} 1 & \text{if } [a,b] \in \{[x,y],[y,z]\} \\ -1 & \text{if } [a,b] = [x,z] \\ 0 & \text{otherwise} \end{cases}$$

- Definition for $\mathbf{B}_k$ with $k \geqslant 2$ is in Appendix.

A coboundary matrix $\mathbf{B}_k^\top$ (adjoint of $\mathbf{B}_k$) maps $(k-1)$-simplex to its k-th cofaces

Remark. $\mathbf{B}_k$ is defined on an SC_ℓ or a CB_ℓ

Boundary matrix

A boundary matrix $\mathbf{B}_k \in \mathbb{R}^{n_k \times n_{k-1}}$ maps a k-simplex to its $(k-1)$-th faces

- With $[x, y, z] \in T$, $\mathbf{B}_1$ and $\mathbf{B}_2$ are defined as:

$$[\mathbf{B}_1]_{a,xy} = \begin{cases} 1 & \text{if } a = x \\ -1 & \text{if } a = y \\ 0 & \text{otherwise} \end{cases} \quad ; \quad [\mathbf{B}_2]_{ab,xyz} = \begin{cases} 1 & \text{if } [a,b] \in \{[x,y],[y,z]\} \\ -1 & \text{if } [a,b] = [x,z] \\ 0 & \text{otherwise} \end{cases}$$

- Definition for $\mathbf{B}_k$ with $k \geqslant 2$ is in Appendix.

A coboundary matrix $\mathbf{B}_k^\top$ (adjoint of $\mathbf{B}_k$) maps $(k-1)$-simplex to its k-th cofaces

Remark. $\mathbf{B}_k$ is defined on an SC_ℓ or a CB_ℓ

Backup slides

Boundary operators

Levi-Civita notation & Permutation parity

Definition S1 (Permutation parity)

Given a finite set $\{j_0, j_1, \cdots, j_k\}$ with $k \geqslant 1$ and $j_\ell < j_m$ if $\ell < m$, the parity of a permutation $\sigma(\{j_0, \cdots, j_k\}) = \{i_0, i_1, \cdots, i_k\}$ is defined to be

$$\epsilon_{i_0,\cdots,i_k} = -1^{N(\sigma)} \tag{S1}$$

Here $N(\sigma)$ is the *inversion number* of σ. The inversion number is the cardinality of the inversion set, i.e., $N(\sigma) = \#\{(\ell, m) : i_\ell > i_m \text{ if } \ell < m\}$. We say σ is an even permutation if $\epsilon_{i_0,\cdots,i_k} = 1$ and an odd permutation otherwise.

Remark. The Levi-Civita symbol for $k = 1$ (left) and 2 (right) is

$$\epsilon_{ij} = \begin{cases} +1 & \text{if } (i,j) = (1,2) \\ -1 & \text{if } (i,j) = (2,1) \end{cases}; \quad \epsilon_{ijk} = \begin{cases} +1 & \text{if } (i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\} \\ -1 & \text{if } (i,j,k) \in \{(3,2,1),(1,3,2),(2,1,3)\} \end{cases}$$

Boundary map for k-cochain

Definition S2 (Boundary map & boundary matrix)

Let $i_0 \cdots \hat{i_j} \cdots i_k \coloneqq i_0, \cdots, i_{j-1}, i_{j+1}, \cdots, i_k$, and $i_0 \cdots \check{i_j} \cdots i_k$ denote i_j insert into $i_0, \cdots, i_k$ with proper order, one can define a *boundary map (operator)* $\mathcal{B}_k : \mathcal{C}_k \to \mathcal{C}_{k-1}$ which maps a simplex to its face by

$$\mathcal{B}_k([i_0, \cdots, i_k]) = \sum_{j=0}^{k} (-1)^j [i_0 \cdots \hat{i_j} \cdots i_k] = \sum_{j=0}^{k} \epsilon_{i_j, i_0 \cdots \hat{i_j} \cdots i_k} [i_0 \cdots \hat{i_j} \cdots i_k] \tag{S2}$$

The corresponding *boundary matrix* $\mathbf{B}_k \in \{0, \pm 1\}^{n_{k-1} \times n_k}$ can be defined as follow

$$(\mathbf{B}_k)_{\sigma_{k-1}, \sigma_k} \begin{cases} \epsilon_{i_j, i_0 \cdots \hat{i_j} \cdots i_k} & \text{if } \sigma_k = [i_0, \cdots, i_k],\ \sigma_{k-1} = [i_0 \cdots \hat{i_j} \cdots i_k] \\ 0 & \text{otherwise.} \end{cases} \tag{S3}$$

$(\mathbf{B}_k)_{\sigma_{k-1}, \sigma_k}$ represents the orientation of σ_{k-1} as a face of σ_k, or equals 0 when the two are not adjacent.

Backup slides

Additional definitions

Neighborhood graphs

Definition S3 (Neighborhood graphs)

δ-*radius graph*: $E = \{(i,j) \in V^2 : \|\mathbf{x}_i - \mathbf{x}_j\|_2 \leqslant \delta\}$;

k-*NN graph*: $E = \{(i,j) \in V^2 : \|\mathbf{x}_i - \mathbf{x}_j\|_2 \leqslant \max\left(\rho_k(\mathbf{x}_i), \rho_k(\mathbf{x}_j)\right)\}$

δ-*CkNN graph* [Berry and Sauer, 2019]: $E = \left\{(i,j) \in V^2 : \frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2}{\sqrt{\rho_k(\mathbf{x}_i)\rho_k(\mathbf{x}_j)}} \leqslant \delta\right\}.$

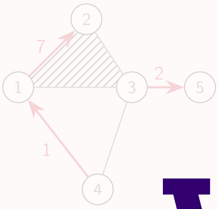
k-COCHAIN

A flow (k-cochain) $\boldsymbol{\omega}_k$ on an SC/CB can be described by a linear combination of k-simplices:

- $\boldsymbol{\omega}_k = \sum_i \omega_{k,i}\sigma_i^k$, where $\sigma_i^k \in \Sigma_k$
- Can further denote by $\boldsymbol{\omega}_k = (\omega_{k,1}, \cdots, \omega_{k,n_k})^\top \in \mathbb{R}^{n_k}$
- Space of $\boldsymbol{\omega}_k$ is $\mathcal{C}_k$, which is isomorphic to $\mathbb{R}^{n_k}$

Example. The flow on the toy SC_2 is $\boldsymbol{\omega}_1 = 7 \cdot [1, 2] + 2 \cdot [3, 5] + (-1) \cdot [1, 4]$, or

$$\boldsymbol{\omega}_1 = \begin{bmatrix} 7 & 0 & -1 & 0 & 0 & 2 \\ [1,2] & [1,3] & [1,4] & [2,3] & [3,4] & [3,5] \end{bmatrix} \in \mathbb{R}^6$$



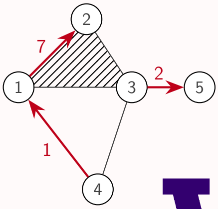
k-COCHAIN

A flow (k-cochain) $\boldsymbol{\omega}_k$ on an SC/CB can be described by a linear combination of k-simplices:

- $\boldsymbol{\omega}_k = \sum_i \omega_{k,i}\sigma_i^k$, where $\sigma_i^k \in \Sigma_k$
- Can further denote by $\boldsymbol{\omega}_k = (\omega_{k,1}, \cdots, \omega_{k,n_k})^\top \in \mathbb{R}^{n_k}$
- Space of $\boldsymbol{\omega}_k$ is $\mathcal{C}_k$, which is isomorphic to $\mathbb{R}^{n_k}$

Example. The flow on the toy SC_2 is $\boldsymbol{\omega}_1 = 7 \cdot [1,2] + 2 \cdot [3,5] + (-1) \cdot [1,4]$, or

$$\boldsymbol{\omega}_1 = \begin{bmatrix} 7 & 0 & -1 & 0 & 0 & 2 \\ [1,2] & [1,3] & [1,4] & [2,3] & [3,4] & [3,5] \end{bmatrix} \in \mathbb{R}^6$$



Gradient, curl, and harmonic components

$$\hat{\mathbf{p}}_0 = \underset{\mathbf{p}_0 \in \mathbb{R}^{n_0}}{\operatorname{argmin}} \|\mathbf{W}_1^{1/2}\mathbf{B}_1^\top \mathbf{p}_0 - \boldsymbol{\omega}\|^2;$$

$$\hat{\mathbf{p}}_2 = \underset{\mathbf{p}_2 \in \mathbb{R}^{n_2}}{\operatorname{argmin}} \|\mathbf{W}_1^{-1/2}\mathbf{B}_2^\top \mathbf{p}_2 - \boldsymbol{\omega}\|^2;$$

$$\hat{\mathbf{h}} = \boldsymbol{\omega} - \underbrace{\mathbf{W}_1^{1/2}\mathbf{B}_1^\top \hat{\mathbf{p}}_0}_{\text{gradient}} - \underbrace{\mathbf{W}_1^{-1/2}\mathbf{B}_2 \hat{\mathbf{p}}_2}_{\text{curl}}.$$

Backup slides

Approximate 1-cochain & underlying vector fields

Linear interpolation of 1-cochain

Let $e = [i, j]$, since $\omega_e = \int_0^1 \zeta(\gamma(t))\gamma'(t)dt$, if given only the vertex-wise vector field $\zeta(x_i) = f(x_i) \in \mathbb{R}^D$, one can approximate the geodesic by $\gamma(t) \approx x_i + (x_j - x_i)t$ and the vector field along γ by $f(\gamma(t)) \approx f(x_i) + (f(x_j) - f(x_i))t$, one has,

$$
\begin{aligned}
\omega_e &= \int_0^1 f^\top(\gamma(t))\gamma'(t)dt \approx \int_0^1 \left[f(x_i) + (f(x_j) - f(x_i))t\right]^\top (x_j - x_i)dt \\
&= \frac{1}{2}(f(x_i) + f(x_j))^\top (x_j - x_i)
\end{aligned}
\tag{S4}
$$

Note that (S4) can be written in a more concise form using *boundary operator* $\mathbf{B}_1$. Let $\mathbf{F} \in \mathbb{R}^{n \times D}$ with $f_i = \mathbf{F}_{i,:} = f(x_i)$. Since $[|\mathbf{B}_1^\top|\mathbf{F}]_{[i,j]} = f(x_i) + f(x_j)$, and $[-\mathbf{B}_1^\top \mathbf{X}]_{[i,j]} = x_j - x_i$. Therefore,

$$\omega = -\frac{1}{2}\operatorname{diag}(\mathbf{B}_1^\top \mathbf{X}\mathbf{F}^\top |\mathbf{B}_1|)$$

Linear interpolation of 1-cochain

Let $e = [i, j]$, since $\omega_e = \int_0^1 \zeta(\gamma(t))\gamma'(t)dt$, if given only the vertex-wise vector field $\zeta(x_i) = f(x_i) \in \mathbb{R}^D$, one can approximate the geodesic by $\gamma(t) \approx x_i + (x_j - x_i)t$ and the vector field along γ by $f(\gamma(t)) \approx f(x_i) + (f(x_j) - f(x_i))t$, one has,

$$\begin{aligned}\omega_e &= \int_0^1 f^\top(\gamma(t))\gamma'(t)dt \approx \int_0^1 \left[f(x_i) + (f(x_j) - f(x_i))t\right]^\top (x_j - x_i)dt \\ &= \frac{1}{2}(f(x_i) + f(x_j))^\top (x_j - x_i)\end{aligned} \tag{S4}$$

Note that (S4) can be written in a more concise form using *boundary operator* $\mathbf{B}_1$. Let $F \in \mathbb{R}^{n\times D}$ with $f_i = F_{i,:} = f(x_i)$. Since $[|\mathbf{B}_1^\top|F]_{[i,j]} = f(x_i) + f(x_j)$, and $[-\mathbf{B}_1^\top X]_{[i,j]} = x_j - x_i$. Therefore,

$$\omega = -\frac{1}{2}\operatorname{diag}(\mathbf{B}_1^\top X F^\top |\mathbf{B}_1|)$$

OBTAINING VERTEX-WISE VECTOR FIELD FROM 1-COCHAIN

Let $\mathbf{X}_E = -\mathbf{B}_1^\top \mathbf{X}$ (so $[\mathbf{X}_E]_{[i,j]} = \mathbf{x}_j - \mathbf{x}_i$) and define $\boldsymbol{\chi}_E$ such that $[\boldsymbol{\chi}_E]_{[i,j]} = \|\mathbf{x}_j - \mathbf{x}_i\|_2^2$. Given the 1-cochain $\boldsymbol{\omega}$, one can solve the following D least square problems to estimate the vector field $\mathbf{F}$ on each point $\mathbf{x}_i$.

$$\hat{\mathbf{v}}_\ell = \underset{\mathbf{v}_\ell \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \left\| |\mathbf{B}_1^\top| \mathbf{v}_\ell - ([\mathbf{X}_E]_{:,\ell} \oslash \boldsymbol{\chi}_E) \circ \boldsymbol{\omega} \right\|_2^2 \right\} \; \forall \ell = 1, \cdots, D \tag{S5}$$

$\circ$, $\oslash$ is *Hadamard product* and *division*, respectively. The solution to the ℓ-th least square problem corresponds to estimate $f_\ell(x_i)$ from $\frac{1}{2}(f_\ell(x_i) + f_\ell(x_j))$. I.e., (inner product)

$$\frac{1}{2}(f_\ell^{\|}(x_i) + f_\ell^{\|}(x_j)) = [([\mathbf{X}_E]_{:,\ell} \oslash \boldsymbol{\chi}_E) \circ \boldsymbol{\omega}]_{[i,j]} = \frac{(x_{j,\ell} - x_{i,\ell})\omega_{ij}}{\|x_j - x_i\|^2}$$

The estimated vector field $\hat{F}$ is

$$\hat{F} = \begin{bmatrix} | & | & & | \\ \hat{v}_1 & \hat{v}_2 & \ldots & \hat{v}_D \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times D}$$

Obtaining vertex-wise vector field from 1-cochain

Let $\mathbf{X}_E = -\mathbf{B}_1^\top \mathbf{X}$ (so $[\mathbf{X}_E]_{[i,j]} = \mathbf{x}_j - \mathbf{x}_i$) and define χ_E such that $[\chi_E]_{[i,j]} = \|\mathbf{x}_j - \mathbf{x}_i\|_2^2$. Given the 1-cochain $\boldsymbol{\omega}$, one can solve the following D least square problems to estimate the vector field $\mathbf{F}$ on each point $\mathbf{x}_i$.

$$\hat{\mathbf{v}}_\ell = \underset{\mathbf{v}_\ell \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \left\| |\mathbf{B}_1^\top| \mathbf{v}_\ell - ([\mathbf{X}_E]_{:,\ell} \oslash \chi_E) \circ \boldsymbol{\omega} \right\|_2^2 \right\} \; \forall \ell = 1, \cdots, D \tag{S5}$$

$\circ$, $\oslash$ is *Hadamard product* and *division*, respectively. The solution to the ℓ-th least square problem corresponds to estimate $f_\ell(\mathbf{x}_i)$ from $\frac{1}{2}(f_\ell(\mathbf{x}_i) + f_\ell(\mathbf{x}_j))$. I.e., (inner product)

$$\frac{1}{2}(f_\ell^{\|}(\mathbf{x}_i) + f_\ell^{\|}(\mathbf{x}_j)) = [([\mathbf{X}_E]_{:,\ell} \oslash \chi_E) \circ \boldsymbol{\omega}]_{[i,j]} = \frac{(x_{j,\ell} - x_{i,\ell})\omega_{ij}}{\|\mathbf{x}_j - \mathbf{x}_i\|^2}$$

The estimated vector field $\hat{F}$ is

$$\hat{F} = \begin{bmatrix} | & | & & | \\ \hat{v}_1 & \hat{v}_2 & \ldots & \hat{v}_D \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times D}$$

Obtaining vertex-wise vector field from 1-cochain

Let $\mathbf{X}_E = -\mathbf{B}_1^\top \mathbf{X}$ (so $[\mathbf{X}_E]_{[i,j]} = \mathbf{x}_j - \mathbf{x}_i$) and define $\boldsymbol{\chi}_E$ such that $[\boldsymbol{\chi}_E]_{[i,j]} = \|\mathbf{x}_j - \mathbf{x}_i\|_2^2$. Given the 1-cochain $\boldsymbol{\omega}$, one can solve the following D least square problems to estimate the vector field $\mathbf{F}$ on each point $\mathbf{x}_i$.

$$\hat{\mathbf{v}}_\ell = \operatorname*{argmin}_{\mathbf{v}_\ell \in \mathbb{R}^n} \left\{ \left\| |\mathbf{B}_1^\top| \mathbf{v}_\ell - ([\mathbf{X}_E]_{:,\ell} \oslash \boldsymbol{\chi}_E) \circ \boldsymbol{\omega} \right\|_2^2 \right\} \; \forall \ell = 1, \cdots, D \qquad \text{(S5)}$$

$\circ$, $\oslash$ is *Hadamard product* and *division*, respectively. The solution to the ℓ-th least square problem corresponds to estimate $f_\ell(\mathbf{x}_i)$ from $\frac{1}{2}(f_\ell(\mathbf{x}_i) + f_\ell(\mathbf{x}_j))$. I.e., (inner product)

$$\frac{1}{2}(f_\ell^{\|}(\mathbf{x}_i) + f_\ell^{\|}(\mathbf{x}_j)) = [([\mathbf{X}_E]_{:,\ell} \oslash \boldsymbol{\chi}_E) \circ \boldsymbol{\omega}]_{[i,j]} = \frac{(x_{j,\ell} - x_{i,\ell})\omega_{ij}}{\|\mathbf{x}_j - \mathbf{x}_i\|^2}$$

The estimated vector field $\hat{F}$ is

$$\hat{F} = \begin{bmatrix} | & | & & | \\ \hat{\mathbf{v}}_1 & \hat{\mathbf{v}}_2 & \dots & \hat{\mathbf{v}}_D \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times D}$$

Backup slides

Applications

Homologous loop detection—theory

Proposition S4 (Induced digraph from z_i)

Let z_i for $i = 1, \cdots, \beta_1$ be the i-th homology basis that corresponds to the i-th homology class and G_i be the induced digraph of the flow z_i. Then for every $i = 1, \cdots, \beta_1$,

1. *there exist at least one cycle in the digraph G_i such that every vertex $v \in V$ can traverse back to itself (reachable);*
2. *the corresponding cycle will enclose at least one homology class (no short-circuiting).*

Sketch of proof.

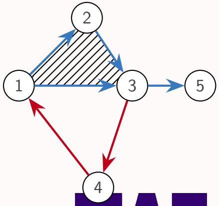
- **Reachable**: harmonic flow is *divergence-free*
- **no short-circuiting**: from Stoke's theorem and Poincaré Lemma [Lee, 2013]

∎

Example.

$$z_i = \begin{bmatrix} 7 & [1,2] \\ 2 & [1,3] \\ -1 & [1,4] \\ 3 & [2,3] \\ -5 & [3,4] \\ 2 & [3,5] \end{bmatrix} \in \mathbb{R}^6$$

$$\Downarrow$$

$G_i =$ 

SPECTRAL HOMOLOGOUS LOOP DETECTION FROM Z

Algorithm S1: SpectralLoopFind

Input: $Z = [z_1, \cdots, z_{\beta_1}]$, V, E, edge distance $\mathbf{d}$

for $i = 1, \cdots, \beta_1$ do

- $E_i^+ \leftarrow \{(s,t) : (s,t) \in E \text{ and } [z_i]_{(s,t)} > 0\}$
- $E_i^- \leftarrow \{(t,s) : (s,t) \in E \text{ and } [z_i]_{(s,t)} < 0\}$
- $\tau \leftarrow \text{Percentile}(|z_i|, 1 - 1/\beta_1)$
- $E_i^\times \leftarrow \{e \in E_i^+ \cup E_i^- : |[z_i]_e| < \tau\}$
- $E_i \leftarrow E_i^+ \cup E_i^- \setminus E_i^\times$
- $G_i \leftarrow (V, E_i)$, with weight of $e \in E_i$ being $[\mathbf{d}]_e$
- $d_{\min} = \texttt{inf}$
- for $e = (t, s_0) \in E_i$ do
  - $\mathcal{P}^* (:= [s_0, s_1, \cdots, t]), d^* \leftarrow \text{Dijkstra}(G_i, \texttt{from=}s_0, \texttt{to=}t)$
  - if $d^* < d_{\min}$ then
    - $\mathcal{C}_i \leftarrow [t, s_0, s_1, \cdots, t]$

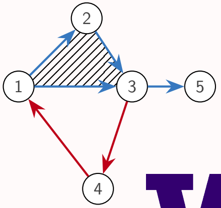
Return: $\mathcal{C}_1, \cdots, \mathcal{C}_{\beta_1}$

Build induced digraph from z_i:

$$z_i = \begin{bmatrix} 7 & [1,2] \\ 2 & [1,3] \\ -1 & [1,4] \\ 3 & [2,3] \\ -5 & [3,4] \\ 2 & [3,5] \end{bmatrix} \in \mathbb{R}^6$$

$$\Downarrow$$

$G_i =$



Spectral homologous loop detection from Z

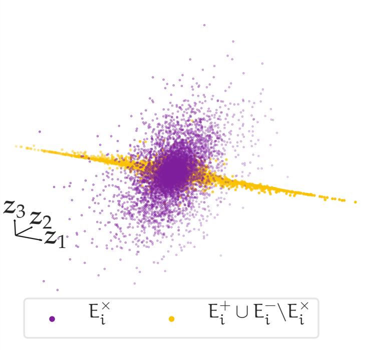
Algorithm S1: SpectralLoopFind

Input: $Z = [z_1, \cdots, z_{\beta_1}]$, V, E, edge distance $\mathbf{d}$

for $i = 1, \cdots, \beta_1$ do
- $E_i^+ \leftarrow \{(s,t) : (s,t) \in E \text{ and } [z_i]_{(s,t)} > 0\}$
- $E_i^- \leftarrow \{(t,s) : (s,t) \in E \text{ and } [z_i]_{(s,t)} < 0\}$
- $\tau \leftarrow \text{Percentile}(|z_i|, 1 - 1/\beta_1)$
- $E_i^\times \leftarrow \{e \in E_i^+ \cup E_i^- : |[z_i]_e| < \tau\}$
- $E_i \leftarrow E_i^+ \cup E_i^- \setminus E_i^\times$
- $G_i \leftarrow (V, E_i)$, with weight of $e \in E_i$ being $[\mathbf{d}]_e$
- $d_{\min} = \texttt{inf}$
- for $e = (t, s_0) \in E_i$ do
  - $\mathcal{P}^* (:= [s_0, s_1, \cdots, t]), d^* \leftarrow \text{Dijkstra}(G_i, \texttt{from=}s_0, \texttt{to=}t)$
  - if $d^* < d_{\min}$ then
    - $\mathcal{C}_i \leftarrow [t, s_0, s_1, \cdots, t]$

Return: $\mathcal{C}_1, \cdots, \mathcal{C}_{\beta_1}$

Thresholding z_i:



- Each homology class has $\approx n_1/\beta_1$ edges

SPECTRAL HOMOLOGOUS LOOP DETECTION FROM Z

Algorithm S1: SpectralLoopFind

Input: $Z = [z_1, \cdots, z_{\beta_1}]$, V, E, edge distance $\mathbf{d}$

for $i = 1, \cdots, \beta_1$ **do**

  $E_i^+ \leftarrow \{(s,t) : (s,t) \in E \text{ and } [z_i]_{(s,t)} > 0\}$

  $E_i^- \leftarrow \{(t,s) : (s,t) \in E \text{ and } [z_i]_{(s,t)} < 0\}$

  $\tau \leftarrow \text{Percentile}(|z_i|, 1 - 1/\beta_1)$

  $E_i^\times \leftarrow \{e \in E_i^+ \cup E_i^- : |[z_i]_e| < \tau\}$

  $E_i \leftarrow E_i^+ \cup E_i^- \setminus E_i^\times$

  $G_i \leftarrow (V, E_i)$, with weight of $e \in E_i$ being $[\mathbf{d}]_e$

  $d_{\min} = \texttt{inf}$

  for $e = (t, s_0) \in E_i$ **do**

    $\mathcal{P}^* (:= [s_0, s_1, \cdots, t]), d^* \leftarrow \text{Dijkstra}(G_i, \texttt{from=}s_0, \texttt{to=}t)$

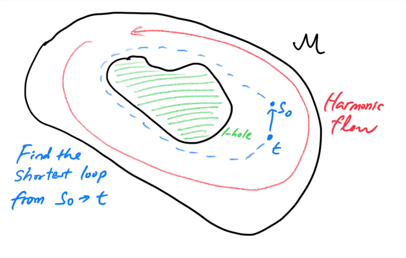
    if $d^* < d_{\min}$ **then**

      $\mathcal{C}_i \leftarrow [t, s_0, s_1, \cdots, t]$

Return: $\mathcal{C}_1, \cdots, \mathcal{C}_{\beta_1}$

Shortest "loop" with Dijkstra:

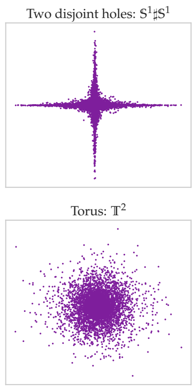
- Dijkstra will find a loop for every $v \in V$ (reachable)
- Every loop obtained is valid (no short-circuiting)



Classifying any 2-dimensional manifold

$\beta_1(\text{torus}) = \beta_1(\text{two disjoint holes}) = 2$

- Not possible to distinguish these two manifolds *only* by rank information
- From Theorem 1, the embedding of $\mathbb{S}^1 \sharp \mathbb{S}^1$ can be (roughly) factorized into two "lines"
- Any loop in $\mathbb{T}^2$ is a convex combination of the two homology classes
  - Intrinsic dimension $= 2$



Remark. Can categorize the manifold $\mathcal{M}$ from $\mathbf{Z}$

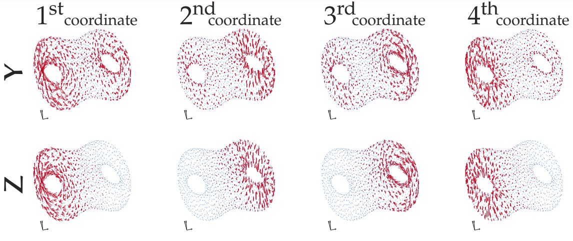
- With the classification theorem of surfaces [Armstrong, 2013]

Proposition S5 (Shape of the embedding $\mathbf{Z}$ of a flat m-torus $\mathbb{T}^m$)

The envelope of the first homology embedding (1-cochain) induced by the harmonic 1-form on the flat m-torus $\mathbb{T}^m$ is an m-dimensional ellipsoid.

Other applications

Visualize the basis of harmonic vector fields:



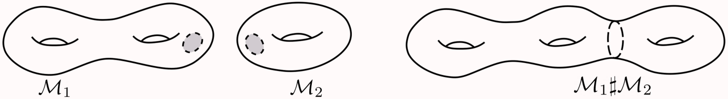
Higher-order simplex clustering [Ebli and Spreemann, 2019]:

- Theorem 1 supports the use of subspace clustering algorithm in this framework

Backup slides

Assumptions and theorems

Notations



	Disjoint manifold $\mathcal{M}_i$	Connected manifold $\mathcal{M}$
Simplicial complex	$\hat{SC}_k^{(i)} = (\hat{\Sigma}_0^{(i)}, \cdots, \hat{\Sigma}_k^{(i)})$	$SC_k = (\Sigma_0, \cdots, \Sigma_k)$
k-Laplacian	$\hat{\mathcal{L}}_k^{(ii)}$	$\mathcal{L}_k$
Homology space	$\mathcal{H}_k(\mathcal{M}_i)$	$\mathcal{H}_k(\mathcal{M})$
k-th Betti number	$\beta_k(\mathcal{M}_i)$	$\beta_k(\mathcal{M})$
Homology embedding	$\hat{\mathbf{Y}}$	$\mathbf{Y}$

Remark.

- Notation with $\hat{\square}$:= disjoint manifolds
- $\hat{SC} = \bigcup_{i=1}^{K} \hat{SC}^{(i)} \neq SC$

Data sampled from a decomposible manifold

Assumption 1

1. *$\mathcal{H}_k(\mathsf{SC})$ (discrete) is isomorphic to the homology group $H_k(\mathcal{M}, \mathbb{R})$ (continuous)*
2. *Assume that $\mathcal{M} = \mathcal{M}_1 \sharp \cdots \sharp \mathcal{M}_\kappa$ and the isomorphic condition holds for every $\mathcal{M}_i$, i.e.,*

$$\mathcal{H}_k(\hat{\mathsf{SC}}^{(i)}) \cong \mathcal{H}_k(\mathcal{M}_i) \text{ for } i = 1, \cdots, \kappa.$$

Remark.

1. Any procedure for constructing SC or weight function for $\mathcal{L}_k$ is acceptable
2. Manifold $\mathcal{M}$ can be decomposed
   - Mostly true except for the known hard case of 4-manifolds

Topology is preserved during connected sum

Assumption 2

Denote the set of destroyed and created k*-simplexes during connected sum by* $\mathfrak{D}_k$ *and* $\mathfrak{C}_k$*;* $\mathfrak{N}_k = \Sigma_k \backslash \mathfrak{C}_k = \hat{\Sigma}_k \backslash \mathfrak{D}_k$ *is the set of non-intersecting simplexes. Then*

1. *no* k*-homology class is created during the connected sum process, i.e.,*

$$\beta_k(\mathsf{SC}) = \sum_{i=1}^{\kappa} \beta_k(\hat{\mathsf{SC}}^{(i)}); \text{ and}$$

2. *The minimum eigenvalues of* $\mathcal{L}_k^{\mathfrak{C},\mathfrak{C}}$ *and* $\hat{\mathcal{L}}_k^{\mathfrak{D},\mathfrak{D}}$ *are bounded away from the eigengaps* δ_i *of* $\mathcal{L}_k^{(ii)}$*, i.e.,* $\min\{\lambda_{\min}(\mathcal{L}_k^{\mathfrak{C},\mathfrak{C}}), \lambda_{\min}(\hat{\mathcal{L}}_k^{\mathfrak{D},\mathfrak{D}})\} \gg \min\{\delta_1, \cdots, \delta_\kappa\}$.

Remark.

1. If $\dim(\mathcal{M}) > k$, then $\mathcal{H}_k(\mathcal{M}_1 \sharp \mathcal{M}_2) \cong \mathcal{H}_k(\mathcal{M}_1) \oplus \mathcal{H}_k(\mathcal{M}_2)$ [Lee, 2013]
2. E.g., it happens when $\mathfrak{C}_k$ and $\mathfrak{D}_k$ are cliques contained in small balls

Small perturbations in the $(k+1)$-simplex set

Assumption 3 (Informal, see also Assumption 6.4 in the thesis)

Let $\tilde{w}_k = |\mathbf{B}_{k+1}[\mathfrak{N}_k, \mathfrak{N}_{k+1}]| w_{k+1}$, $\tilde{w}_{k-1} = |\mathbf{B}_k[:, \mathfrak{N}_k]| \tilde{w}_k$. *For* $\ell = k$ *or* $k-1$, *we have*

$\lvert\mathfrak{C}_k\rvert$ *is small:*	$\max_{\sigma\in\mathfrak{N}_\ell} \{w_\ell(\sigma)/\tilde{w}_\ell(\sigma) - 1\} \leqslant \epsilon_\ell;$
$\lvert\mathfrak{D}_k\rvert$ *is small:*	$\max_{\sigma\in\mathfrak{N}_\ell} \{\hat{w}_\ell(\sigma)/\tilde{w}_\ell(\sigma) - 1\} \leqslant \epsilon_\ell;$ *and*
The **net** *effect is small:*	$\max_{\sigma\in\mathfrak{N}_\ell} \{\lvert w_\ell(\sigma)/\hat{w}_\ell(\sigma) - 1\rvert\} \leqslant \epsilon'_\ell.$

1. Not too many triangles are created/destroyed during connected sum
2. Sparsely connected manifold
   - Density in the connected sum region should be smaller than other regions
3. Empirically, the perturbation is small even when $\mathcal{M}$ is not sparsely connected

Subspace perturbation: sketch of proof of Theorem 1

Sketch of proof. The proof (in Supplement) is based on

1. Bound the error ($\mathsf{DiffL}_k^{\mathsf{up}}$ and $\mathsf{DiffL}_k^{\mathsf{down}}$ terms) between $\mathcal{L}_k$ and $\hat{\mathcal{L}}_k$ with $\tilde{\mathcal{L}}_k$;
   - $\tilde{\mathcal{L}}_k :=$ the Laplacian after removing the k-simplices in both $\mathfrak{C}_k$ and $\mathfrak{D}_k$ during connected sum
2. Use of a variant of the Davis-Kahan theorem [Yu et al., 2015] (the spectral norm $\|\cdot\|$); and
3. Bound the spectral norm of $\mathcal{L}_k$ for a simplicial complex [Horak and Jost, 2013]

$$\|\mathcal{L}_k\|_2 \leqslant k+2.$$

   - Any $(k+1)$-simplex has $(k+2)$ faces

Subspace perturbation for cubical complex

Proposition S6

Given an up $\mathbf{k}$*-Laplacian* $\mathcal{L}_k^{\mathrm{up}} = \mathbf{A}_{k+1}\mathbf{A}_{k+1}^\top$ *with* $\mathbf{A}_{k+1} = \mathbf{W}_k^{-1/2}\mathbf{B}_{k+1}\mathbf{W}_{k+1}^{1/2}$ *built from a cubical complex, we have*

$$\|\mathcal{L}_k^{\mathrm{up}}\|_2 \leqslant \lambda_k = 2k+2.$$

Sketch of proof. The $(2k+2)$ term comes from the fact that a $(k+1)$-cube has $(2k+2)$ faces. The rest of the proof follows from [Horak and Jost, 2013]. ■

Corollary S7 ($\mathcal{L}_k$ built from a cubical complex)

Under Assumptions 2–3 with $\mathrm{DiffL}_k^{\mathrm{up}}$ *as well as* $\mathrm{DiffL}_k^{\mathrm{down}}$ *defined in Theorem 1 and* $\lambda_k = 2k+2$*, there exists a unitary matrix* $\mathbf{O}$ *such that (1) holds.*